DOCUMENT RESUME

ED 399 272 TM 025 428

AUTHOR Lietz, Petra H.; Roche, Lawrence A.
TITLE Testing the Invariance of Reading Literacy Dimensions across Different Countries: An Application of Multigroup Confirmatory Factor Analysis.
PUB DATE Apr 96
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adolescents; *Correlation; Databases; Factor Analysis; *Factor Structure; Foreign Countries; Goodness of Fit; *International Studies; *Literacy; Mathematical Models; *Reading Comprehension; Surveys; Test Items
IDENTIFIERS Belgium; Hungary; *Invariance; Italy; United States

ABSTRACT

This study investigates whether or not the factor structure of reading comprehension is invariant across large, nationally representative samples of 14-year-old students from four different countries. The data from French-speaking Belgium, Hungary, Italy, and the United States were collected as part of the Reading Literacy Study of 1990-91, conducted by the International Association for the Evaluation of Educational Achievement. The relevance and application of multigroup confirmatory factor analysis techniques (K. G. Joreskog and D. Sorbom, 1988) to the assessment of model generalizability across countries or cultures, particularly in relation to international databases, is demonstrated and discussed. Results indicate that it is not unreasonable to assume factor loadings and factor correlations to be invariant across the four countries. When item uniquenesses were set to be invariant, a decrease in fit according to the relative noncentrality index and the Tucker Lewis index was observed. In contrast, the parsimony relative noncentrality index showed a higher value for the model in which invariance was assumed for all parameters in the model. (Contains 1 table, 1 figure, and 27 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 399 272

Testing the Invariance of Reading Literacy Dimensions Across Different Countries: An Application of Multigroup Confirmatory Factor Analysis.

Petra H. Lietz
School of Education, The Flinders University of South Australia

Lawrence A. Roche
Faculty of Education, University of Western Sydney, Macarthur

Paper presented at the American Educational Research Association (AERA) annual meeting, New York, April 8-12, 1996.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
PETRA H. LIETZ
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Requests for reprints or other correspondence may be directed to either author:
Petra Lietz, School of Education, The Flinders University of South Australia, GPO Box 2100, Adelaide, S.A. 5001, e-mail: edpel@cc.flinders.edu.au
GAMGEE.

Lawrence Roche, Faculty of Education, University of Western Sydney, Macarthur, PO Box 555, Campbelltown, N.S.W. 2560, e-mail: l.roche@uws.edu.au

TM 025428

BEST COPY AVAILABLE

Abstract

This study investigates whether or not the factor structure of reading comprehension is invariant across large, nationally representative samples of 14-year-old students from four different countries. The data was collected as part of the Reading Literacy of 1990/91, conducted by the International Association for the Evaluation of Educational Achievement (IEA). The relevance and application of multigroup confirmatory factor analysis techniques (Jöreskog and Sörbom, 1988) to the assessment of model generalizability across countries or cultures -- particularly in relation to international databases -- is demonstrated and discussed.

Objectives

Educational researchers and practitioners interested in cross-cultural comparisons of empirical findings have historically relied on relatively crude analytical tools to assess the applicability of instruments and models to different cultural settings. Today, there is growing awareness of the need to employ more appropriate techniques for testing the generalizability of constructs and their valid measurement across countries or cultures.

This paper aims to examine the generalizability of the factor structure underlying reading comprehension across four different cultures. The models to be examined are based on the test design for the Reading Literacy Study (Elley, 1994) conducted by the International Association for the Evaluation of Educational Achievement (IEA). Previous research using this data has examined alternative models of reading comprehension (Balke, 1995; Gustafsson, 1995). These analyses, however, were restricted to data from the Nordic countries (Balke, 1995), or sought mainly to provide evidence for a juxtaposed general factor (Gustafsson, 1995). More generally, inadequate attention has been paid to the issue of potential cross-cultural differences in the factor structure of reading comprehension. This concern is not about differences in the mean level of reading comprehension, or the basic psychometric adequacy of items across cultures. Rather, the central question is whether or not items within each of the scales in a test measure the same component subskills for different cultural groups. Multigroup comparisons using confirmatory factor analysis (CFA; Jöreskog and Sörbom, 1988) provide a powerful test of alternative models in which specific parameter estimates, sets

of parameter estimates, or all parameter estimates can be constrained to be invariant across groups (Bollen, 1989; Marsh, 1994). This study investigates whether or not a model in which items are assigned to only one of three correlated factors can be shown to be invariant across four countries, namely Belgium (French), Hungary, Italy, and the United States of America.

Theoretical Framework

Reading comprehension is usually regarded as that aspect of reading which allows readers to react towards and make judgements about what they have read and incorporate the new information into their mental concepts (Pearson and Johnson, 1972). Many researchers have attempted to examine the operations involved in reading comprehension and whether or not distinct subskills might be identified or ordered hierarchically. While some research has produced evidence for the existence of separate skills (Davis, 1968; Spearritt, 1972) other researchers have found no support for the existence of multiple dimensions in reading comprehension (Thorndike, 1973; Zwick, 1987). Whereas a hierarchical ordering of reading skills was assumed in the rationale for the objectives to be assessed in the reading tests of the National Assessment of Educational Progress in 1970/71, the information on the objectives for the 1983/84 tests explicitly stated that no such hierarchy could be anticipated. In the context of the Reading Literacy Study, great care was taken to design a test measuring three different domains of reading, namely Narrative, Expository and Documents. These domains were defined as:

(1) Narrative prose: Continuous text in which the writer's aim is to tell a story - whether fact or fiction. They normally follow a linear time sequence and are usually intended to entertain or involve the reader emotionally.

(2) Expository prose: Continuous text designed to describe, explain, or otherwise convey factual information or opinion to the reader.

(3) Documents: Structured information displays presented in the form of charts, tables, maps, graphs, lists or sets of instructions. (Elley et al., 1992, p.4)

Elley et al. also noted that it was intended to report test scores separately for each of the three domains. As a result, most of the reporting of bivariate relationships between student reading achievement and certain student, teacher and school variables was undertaken by providing separate figures for each domain. However, in the major summary report of the Reading Literacy Study, Elley et al. (1994 p. 12) mention that measures of student abilities in reading literacy were estimated and reported for each domain separately as well as for the total item scale.

Subsequent item analyses were undertaken using an international pooled dataset and included all items from the different domains except those which, according to the authors, had poor psychometric properties on the international scale. Again, some results were reported for the total item scale, suggesting a certain ambiguity as to the appropriate reporting of test scores for the Reading Literacy Study, which no attempt would appear to have been made to resolve.

In response to the apparent uncertainty surrounding the factor structure of reading comprehension responses, a number of studies have used CFA to investigate this issue with the IEA data. From the assumptions underlying the test design of the Reading Literacy Study as well as previous research (Balke, 1995; Gustafsson, 1995; Lietz, 1994) several models of the structure underlying reading comprehension are tested for invariance across countries in this paper. Of primary interest is whether or not a correlated three-factor model can be shown to be invariant across the four countries under review in this study, namely Belgium (French), Hungary, Italy and the United States of America.

Method

Confirmatory factor analysis (CFA) is used to evaluate the fit of an a priori model to the data collected (Jöreskog and Sörbom, 1988). Goodness of fit indices are used to assess how closely a matrix reproduced from parameter estimates for the posited model correspond to the input correlation or covariance matrix based on the actual data. A more detailed introduction to the conduct of CFA is available elsewhere (Bollen, 1989; Byrne, 1989; Jöreskog and Sörbom, 1988), and instructive examples of the application of CFA to the issue of factorial invariance across different populations are becoming more common in educational and psychological research (e.g., Marsh, 1993; Marsh & Roche, in press; McInerney, Roche & McInerney, 1994).

The relevant parameters in typical CFA studies consist of factor loadings (relations between measured variables and latent factors); factor variances and covariances (relations among the factors); and item uniquenesses (a combination of specific and error variance). In order to test the invariance of a hypothesised structure across groups, it is necessary to begin with a model that fits the data well (Bentler, 1990; Marsh, 1994). The generalizability of that model across different populations is then evaluated by testing alternative models in which specific parameter estimates (such as factor loadings for selected items), sets of parameter estimates (such as all factor loadings) or all parameter estimates (factor loadings, factor correlations and factor uniquenesses) can be constrained to be invariant (that is, forced to be equivalent) across groups. Invariance in relation to factor loadings is a minimal criterion in multigroup comparisons, but it is also desirable to assess the equivalence of factor correlations and item uniquenesses (Marsh, 1994).

In assessing the fit of a model, it is important to establish firstly that the model converges to a proper solution (eg., no impossible parameter estimates); that the parameter estimates "make sense" in relation to the a priori model and common sense; and finally to evaluate different fit indices in relation to rules of thumb and values from alternative models (Marsh, 1994). Based on evaluations and recommendations of various fit indices (e.g. Marsh & Balla, 1994; Marsh, Balla & Hau, in press), the Tucker Lewis index (TLI) is emphasized, but other indices including the relative noncentrality index (RNI) and its counterpart, the Parsimony RNI (PRNI) which penalizes model complexity - or rewards model parsimony - are also presented. The three indices of fit differ in that the TLI and PRNI provide a control for model parsimony whereas the RNI does not.

These characteristics of different indices are particularly relevant when comparing models with different invariance constraints. As more parameters are constrained to be equal across groups, there are fewer parameters to be estimated, so that the model becomes more parsimonious. Indices such as the chi-square statistic, the Goodness of Fit Index (GFI), and the RNI contain no penalty for lack of parsimony. Thus they are always automatically lower when fewer parameters are estimated, but this may be a result of a reduced likelihood of capitalisation on chance rather than reflecting a less satisfactory model. The penalty for model complexity in the TLI means that it is technically possible for more parsimonious models to obtain a better fit (McDonald & Marsh, 1990). The PRNI imposes a more severe penalty on more complex

models, providing a less conservative test of improvement in fit as the model is constrained to be equivalent between groups.

Data Source

Twenty-two countries participated in the IEA Reading Literacy Study at the 14-year-old level. The four countries for which data are examined in this paper, namely Belgium (French) Hungary, Italy, and the United States of America were chosen to represent relatively distinct cultures within a larger study (Lietz, 1995) which examines changes in reading achievement over time in the eight countries that also participated in the first study of reading comprehension conducted by IEA at the 14-year-old level in 1970/71 (Thorndike, 1973a).

There were 89 core items (mostly multiple-choice as well as some open-ended questions) which were scored for all countries. The questions were based on a total of 19 passages which represented one of three domains, namely narrative prose, expository prose or documents (Elley, 1994). In the final form of the tests, 29 items were assigned to the narrative domain, 26 items to the expository domain and 34 to the document domain.

Data collected from large nationally representative samples of students in Belgium (French), Hungary, Italy, and the United States of America form the evidence available in this study to examine the proposed model of the structure underlying reading comprehension.

Results

Figure 1 illustrates the ways in which different parameters of the basic model (Model A) were constrained to be invariant across groups. As mentioned above, invariance in relation to factor loadings is a minimal criterion in multigroup comparisons. Hence, Model B presents the model in which factor loadings were set to be invariant across the four countries under review. In Model C, correlations between the three factors, Narrative, Expository, and Document were held constant across countries in addition to factor loadings. Model D examined the fit of a structure in which only the correlations between factors were assumed to be invariant. In addition to this constraint, Model E assumed the item uniquenesses to be invariant. Finally, in Model F, total invariance for all parameters across the four groups was examined.

Preliminary analyses were undertaken to eliminate poor items. For this purpose, Model A was examined for each of the four countries separately and items with a factor loading below 0.40 were noted. Where an item showed a low factor loading in three or four countries, it was considered not to represent the underlying domain appropriately and was hence removed. In this way, 15 items were identified and subsequently removed from the analysis, one from the Narrative domain, five from the Expository domain, and nine from the Document domain.

The remaining 74 items were grouped into 22 item parcels representing the means of between two and five items relating to particular passages within each domain. An earlier analysis of the same data based on individual items produced solutions that were suboptimal for testing invariance of parameter estimates (Roche & Lietz, 1995). The use of item parcels is common in factor analytic research (eg., Marsh & Roche, in press), since it results in more valid and reliable indicators, decreasing the effects of idiosyncrasies associated with particular items (particularly in relation to dichotomously scored achievement data). It also reduces the number of measured variables in the model, though the advantage of this in confirmatory factor analysis has not been clearly established (Marsh, Hau & Balla, 1996). Seven item parcels were used as indicators to define the Narrative domain, six parcels were assigned to the Expository domain and nine parcels were assigned to the Document domain. All analyses were undertaken using LISREL8 for Windows (Jöreskog and Sörbom, 1993).

Table 1 presents the goodness-of-fit indices for the confirmatory factor analysis undertaken in this study. In the upper panel of the table, results are reported for the examination of the basic model in which each of the 22 item parcels was assigned to only one of three factors. First, an analysis was undertaken of a data set in which information for the four countries was combined. The goodness-of-fit indices ranged from 0.83 for the PRNI to 0.93 for the RNI, indicating that the model fitted the data quite well. Likewise, the goodness-of-fit indices in Table 1 for the separate analyses of the basic model for the four countries ranged from 0.78 for the PRNI in Belgium (French) to 0.94 for the RNI in Hungary, suggesting the appropriateness of the model. The highest values for goodness-of-fit indices were obtained in Hungary while Belgium (French) showed the relatively lowest values.

Byrne (1989) and Marsh (1993, 1994) recommend to proceed with analyses of invariance after a model has shown an acceptable fit to the data.

Hence, it was decided to test the generalizability of the model across the four countries by undertaking a multigroup confirmatory factor analysis. In this way it could be examined how different assumptions regarding the invariance of different model parameters across the four countries would affect the model.

In the lower panel of Table 1, results are presented of the analysis of each of the models that were illustrated in Figure 1. It should be noted that Model A in which all parameters were allowed to vary across groups showed the best fit according to both the RNI (0.99) and TLI (0.99). In contrast, the PRNI for this model (0.88) was lower than for Models C, E and F due to the lack of parsimony in this model, in which relatively more relationships have to be estimated.

When factor loadings were held invariant in model B, there was a substantial decline in both RNI (.92) and, more significantly - because it rewards the improved parsimony - the TLI (.91). In addition, the PRNI, which provides the most handsome compensation for increased parsimony, also fell slightly. These results suggest that the model allowing factor loadings to differ across countries is a better model. Nevertheless, model B, holding factor loadings invariant, also provides a good fit to the data.

Among models B, C and D, values for the RNI and TLI varied only slightly, suggesting little difference as to whether just the factor loadings, the correlations between factors or both of these parameters together were held constant. This provides some additional support for the relative consistency of the structure across countries, in that relations between the latent variables are similar in each group. Again, however, model A represents a better fit than either models C or D according to the RNI and TLI indices.

Once uniquenesses were incorporated into the assumptions of invariance across groups (Models E and F), values for the RNI and TLI dropped slightly in comparison to models A and B. In contrast, the PRNI improved due to the increased parsimony. Overall, these indices indicate that model F, which posits complete invariance, provides a good fit to the data, particularly when it is noted that conventional rules of thumb (indices over .9 representing good fit) may be particularly conservative when applied to incremental indices such as the RNI and TLI (Hu & Bentler, 1995). In comparison to model A, however, only the heavily pro-parsimony PRNI suggests a better fit for model F, while the generally better-behaved TLI, as well as the RNI, indicates that leaving parameters to be free across groups provides the best fit.

Conclusion

The results presented here indicated that it was not unreasonable to assume that factor loadings and factor correlations be invariant across the four countries. When item uniquenesses were set to be invariant, a decrease in fit according to the RNI and TLI was observed. The PRNI, in contrast, showed the highest value for the model in which invariance was assumed for all parameters in the model.

The overall superiority of the no invariance model, however, suggests that while the parcels used to form indicators in this study (and the passages on which they are based) demonstrate consistently high loadings on their respective factors across all groups, there are detectable variations across the different cultures. Perhaps it is not surprising that different passages would elicit at least some degree of idiosyncratic responses from different cultual groups.

Thus, to the extent that the factor structures were invariant across different countries, this evidence suggested that translation and cultural issues are less of a concern in the assessment of reading comprehension that commonly assumed. The relative consistency found here supported Thorndike's (1973a) conclusion that translation problems could be overcome. Finally, the study demonstrates the importance and utility of multigroup CFA in relation to large, culturally diverse data sets.

References

Balke, G. (1995, April). Decomposition of reading comprehension - Analysis of the IEA literacy tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Bentler, P. M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, 238-246.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Byrne, B. M. (1989). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer Verlag.

Davis, F.B. (1968). Research in comprehension in reading. Reading Research Quarterly, 3, 499-545.

Elley, W.B. (ed.) (1994). The IEA study of reading literacy: Achievement and instruction in thirty-two school systems. Oxford: Pergamon Press.

Elley, W.B., Schleicher, A. and Wagemaker, H. (1994). Introduction. In Elley W. B. (ed.). The IEA study of reading literacy: Achievement and instruction in thirty-two school systems. (pp. 1-34). Oxford: Pergamon Press.

Gustafsson, J.-E. (1995, April). Alternative hierarchical models of reading achievement. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage.

Jöreskog, K.G. & Sörbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago: SPSS, Inc.

Jöreskog, K.G. & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Lietz, P. (1994, November). Issues in the measurement of change in reading achievement over time. Paper presented at the Australian Association for Research in Education Annual Conference, Newcastle, N.S.W.

Lietz, P. (1995). Changes in reading comprehension across cultures and over time. PhD thesis under examination. The Flinders University of South Australia, Adelaide, South Australia.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. American Educational Research Journal, 30, 841-886.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A mutifaceted approach. Structural Equation Modeling, 1, 5-34.

Marsh, H. W., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. Quality and Quantity: International Journal of Methodology, 28,. 185-217.

Marsh, H. W., Balla, J. R., & Hau, K. T. (in press). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In

G. A. Marcoulides & R. E. Schumacker (eds.). Advanced structural equation modeling techniques. Hillsdale, NJ: Erlbaum.

Marsh, H. W., Hau, K. T., & Balla, J. R. (1996). Is more ever too much: The number of indicators per factor in confirmatory factor analysis. Manuscript submitted for publication.

Marsh, H. W., & Roche, L. A. (in press). The structure of artistic self-concept performing arts and non-performing arts students in a performing art school: "Setting the stage" with multigroup confirmatory factor analys of Educational Psychology.

McInerney, D. M., Roche, L. A., & McInerney, V. (1994, November). Uni goals of school motivation?: An application of LISREL to cross-cu research. Paper presented at the Australian Association for Resea Education National Conference, Newcastle, NSW.

Mulaik, S.A. & Quartetti, D.A. (1994). First-order or higher-order? presented to the Society of Multivariate Experimental Psych Annual Meeting, Princeton, New Jersey, October 28. 1994.

Pearson, P.D. & Johnson, D.D. (1972). Teaching reading comprehension. York: Holt,Rinehart and Winston.

Roche, L. A. & Lietz, P. (1995). Where words meet numbers: Testing the generalizability of reading comprehension models across different countries in a demonstration of multigroup confirmatory factor ana Paper presented at the Australian Association for Research in Educ (AARE) annual conference, Hobart, Tasmania.

Spearritt, D. (1972). Identification of subskills in reading comprehension. Rea Research Quarterly, 8 (1), 92-111.

Thorndike, R.L. (1973a). Reading comprehension education in fifteen count Stockholm: Almqvist and Wiksell.

Thorndike, R.L. (1973b). Reading as reasoning. Reading Research Quarter (2) 135-147.

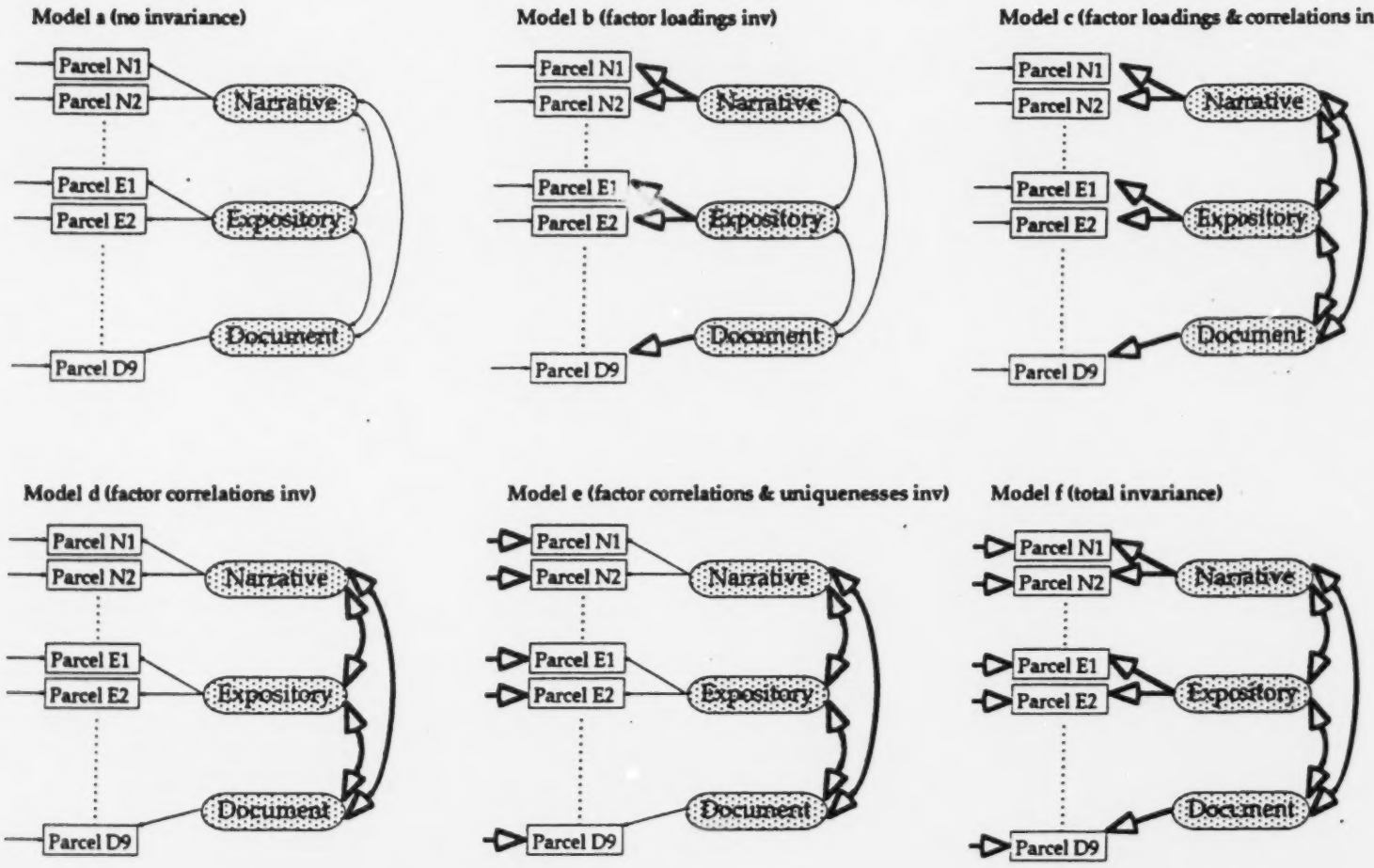
Zwick, R. (1987). Assessment of the dimensionality of NAEP year 15 read data. In A. Beaton (Ed.). The NAEP 1983-84 technical report. (pp. 2 284). Princeton, New Jersey: Educational Testing Service.

Table 1 *Goodness-of-fit indices for confirmatory factor analyses*

Model	N of students	χ^2	df	RNI	TLI	PRNI
Total	12,642	6211.39	206	0.926	0.917	0.825
Belgium (French)	2,732	1930.54	206	0.875	0.859	0.780
Hungary	3,374	1405.12	206	0.941	0.933	0.839
Italy	3,078	1548.28	206	0.925	0.916	0.825
United States	3,458	2377.19	206	0.930	0.922	0.829
A) 4 gp (no inv)		1761.13	824	0.989	0.987	0.882
B) 4 gp (fl inv)		7734.10	881	0.917	0.913	0.875
C) 4 gp (fl, c inv)		8042.14	899	0.914	0.912	0.889
D) 4 gp (c inv)		7490.95	842	0.920	0.912	0.838
E) 4 gp (c, u inv)		8883.67	908	0.904	0.902	0.888
F) 4 gp (tot inv)		9324.10	965	0.899	0.904	0.939

Notes:

χ^2 - Chi-Square
df - Degrees of freedom
RNI - Relative noncentrality index
TLI - Tucker Lewis index
PRNI - Parsimony index for RNI
4 gp (no inv) - four group model with no invariance constraints (Model A)
4 gp (fl inv) - four group model with factor loadings invariant (Model B)
4 gp (fl, c inv) - four group model with factor loadings and factor correlations invariant (Model C)
4 gp (c inv) - four group model with factor correlations invariant (Model D)
4 gp (c, u inv) - four group model with factor correlations and uniquenesses invariant (Model E)
4 gp (tot inv) - four group model with factor loadings, factor correlations and uniquenesses invariant (Model F)



Note: Bold lines represent parameters held invariant between groups.

Figure 1 *Alternative models testing invariance of different sets of parameters*

BEST COPY AVAILABLE